Towards a Fast Steady-State Visual Evoked Potentials (SSVEP) Brain-Computer Interface (BCI)

Aung Aung Phyo Wai[1], **Yangsong Zhang**[2], **Heng Guo**[3], **Ying Chi**[3], **Lei Zhang**[3],
Xian-Sheng Hua[3], **Seong Whan Lee**[4] **and Cuntai Guan**[1, *]

[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore.
[2]School of Computer Science and Technology, Southwest University of Science and Technology, China.
[3]Health-AI Division, DAMO Academy, Alibaba Group Holding Limited.
[4]Department of Brain and Cognitive Engineering, Korea University, South Korea.

February 5, 2020

Abstract

Steady-state visual evoked potentials (SSVEP) brain-computer interface (BCI) provides reliable responses leading to high accuracy and information throughput. But achieving high accuracy typically requires a relatively long time window of one second or more. Various methods were proposed to improve sub-second response accuracy through subject-specific training and calibration. Substantial performance improvements were achieved with tedious calibration and subject-specific training; resulting in user's discomfort. Real-world SSVEP applications demand training-free fast SSVEP frequency recognition with improved performance and usability. So, we propose a training-free method by combining spatial-filtering and temporal alignment (CSTA) to recognize SSVEP responses in sub-second response time. CSTA exploits linear correlation and non-linear similarity between steady-state responses and stimulus templates with complementary fusion to achieve desirable performance improvements. We evaluated the performance of CSTA in terms of accuracy and Information Transfer Rate (ITR) in comparison with both training-based and training-free methods using two SSVEP data-sets. We observed that CSTA achieves the maximum mean accuracy of 97.43 ± 2.26 % and $85.71 \pm 13.41\%$ with four-class and forty-class SSVEP data-sets respectively in sub-second response time in offline analysis. CSTA yields significantly higher mean performance ($p < 0.001$) than training-free method on both data-sets. Compared with training-based methods, CSTA shows 29.33 ± 19.65 % higher mean accuracy with statistical significant differences in time window less than 0.5 s. In longer time windows, CSTA exhibits either better or comparable performance though not statistically significant better than training-based methods. We show that the proposed method brings advantages of subject-independent SSVEP classification without requiring training while enabling high target recognition performance in sub-second response time.

Keywords Steady-State Visual Evoked Potential (SSVEP) · Brain-Computer Interface (BCI) · Sub-second response time · Spatial filtering · Temporal alignment

1 Introduction

Brain-computer interface (BCI) is a promising technology that can enable users to directly interact with target physical or virtual interfaces through brain signals without conventional neural pathways [1]. EEG is widely adopted due to its portability, high temporal resolution, low-cost and high usability. BCI systems are implemented in different paradigms, such as motor imagery [2, 3], Steady-State Visual Evoked Potentials (SSVEP) using periodic frequency stimuli [4, 5], P300 event-related potentials [6, 7], etc., depending on specific application needs. Among these BCI modalities, SSVEP provides high performance throughput, requires less or no training and shows high resilience to interference [1].

---

*Cuntai Guan is the corresponding author. Email: `ctguan@ntu.edu.sg`

SSVEP, in response to repetitive stimulus presentation, can easily be characterised by robust steady-state responses (SSR) with less dependence on time and phase synchronization requirements [8]. So SSVEP-based BCI system has become popular in visual speller and control applications for having reliable responses and high information transfer rate (ITR) [9].

For capturing reliable SSR, SSVEP stimulus presentation generally requires continuous time window of a few seconds over multiple trials [10]. This limits SSVEP in real-time usage as long exposure to repetitive flickering stimuli cause user discomfort leading to performance deterioration and impaired usability [8]. In recent years, several methods were proposed to improve the accuracy while reducing the time window. These new methods leverage correlated features from multi-channel EEG for improving performance in short response time, such as multivariate synchronization index (MSI) [11, 12], task-related component analysis (TRCA) [13], correlated component analysis [14, 15]. Besides the visual speller, SSVEP-BCI could potentially be used to objectively assess the visual field abnormality and its integrity over time in Glaucoma patients [16, 17]. The objective SSVEP responses show better Glaucoma detection than conventional Psycho-physics method that relies on subjective user responses [18]. The presence or absence of SSR in response to multi-frequency modulated stimuli can be regarded as objective measures in discriminating glaucomatous eyes from healthy ones. Recent study using portable SSVEP BCI solution with multiple Glaucoma patients showed high diagnosis accuracy and low test-retest variability [17]. In order to be effective assessment, these clinical assessments require short test duration without tedious calibration stage, and should robustly detect visual functional defects of all subjects. In all SSVEP applications, shorter test duration brings better user experience as well as better test performance because long testing time for visual field assessment is deemed unreliable and lack of user's adherence [8].

The high accuracies in SSVEP decoding results usually rely on long data segment length [11, 19]. For quick responsiveness and online BCI applications, it is important to achieve reliable SSR recognition in short time window, preferably sub-second response time (within 1 s interval). Although standard CCA method is an unsupervised method without requiring calibration or training data, some of its enhanced variants were implemented in supervised way that require training or calibration data [9, 20, 21, 22]. Still, there is poor performance in short time windows compared with longer ones in both supervised and unsupervised variants of CCA solutions [14, 21]. The huge gap in accuracy among different time windows might be due to the poor linear relationship in short segment length [23]. On the other hand, this might be due to non-linear origin of SSVEP responses resulting in SSR with highly non-linear relationship [23]. What's more, collecting calibration data is time-consuming and labor-intensive, and sometimes difficult to obtain representative calibration data. The major issue is that performance differences between sub-second and more than one second response time of most SSVEP methods are highly distinct with less satisfactory performance [24]. Moreover, existing studies only evaluated on single data-set potentially optimizing the methods with data-specific improvement instead of generalizable performance across different data-sets. It is highly desirable having a single method that should be working well on different data-sets with consistence performance across subjects. Nevertheless, various methods were proposed to improve sub-second response accuracy through subject-specific training to learn representative features per stimulus [13, 14]. Those existing methods show superior performance improvements but still a huge discrepancy in mean accuracy between time window less than 0.5 s (sub-second response time) and time window of 1 s or longer. We strongly believe that there is a room for performance improvements in sub-second response time with existing training-free as well as training-based methods [24].

So, we propose an unsupervised method without subject-dependent calibration, termed as combined spatial-filtering and temporal alignment (CSTA), which leverages features from linear correlation and multi-phased non-linear mapping between EEG signals and stimulus reference signals. In speech and human activity recognition tasks, Dynamic Time Warping (DTW) is successfully applied by aligning temporal similarities between two signals to achieve desirable classification or detection performance [25]. DTW method leverages temporal similarity measures to optimize non-linear similarity mapping through distance metrics between measured signals and templates [26]. Some preliminary results with four-class SSVEP data-set showed performance improvements in sub-second response time by combining CCA and DTW [27]. We will evaluate the proposed method in comparison with two training-based and one training-free state-of-the-art (SOTA) methods using two data-sets [20, 21, 13]. The offline evaluation results show significant performance improvements in different short time windows from 0.2 to 1.0 s with CSTA in comparison with other methods used in evaluation [13, 21]. These promising results encourage that non-training method can still achieve similar or better performance compared with those training-based methods; moving towards practical fast SSVEP-based BCI online applications.

In this paper, we propose subject-independent training-free method, CSTA and evaluate its performance against the SOTA methods using two SSVEP data-sets. In Section II, we explain the principles of the proposed method in details. We then explain evaluation scenario and comparison with selected methods using two SSVEP data-sets in Section III. We discuss the details of analysis results using different evaluation criteria in Section IV. Finally, we conclude the paper with possible improvements with some limitations and future works in Section V.

2 Combined Spatial Filtering and Temporal Alignment Method

In order to improve recognition accuracy and reduce user's discomfort or fatigue, SSVEP detection method requires recognizing correct steady-state responses (SSR) in short response time without subject-specific calibration. Our method, CSTA exploits features from spatial filtering and temporal mapping with phase shifts to recognize the target frequency [27]. The spatial filtering with CCA optimizes the 'spatial weights' to find maximum linear correlation between multi-channel EEG signals and reference signals. In contrast, DTW optimizes non-linear 'temporal alignment' matching to find minimum mapping distance measures between EEG signals and stimulus templates (similar reference signals used in CCA).

2.1 CCA Spatial Filtering

CCA is a multivariate analysis method for finding the linear relationship between two multidimensional variables. It optimizes the canonical weight vectors that maximizes the linear relationship between the input variables [28]. For SSVEP recognition, only two multivariate inputs of EEG multi-dimensional signals, $\boldsymbol{X} \in \mathbb{R}^{N_c \times N_s}$ and reference signals, $\boldsymbol{Y} \in \mathbb{R}^{2N_h \times N_s}$ are required; where N_c is number of channels, N_s is number of data samples and N_h is number of harmonics. CCA optimizes two weight vectors $\boldsymbol{w_x} \in \mathbb{R}^{N_c \times 1}$ and $\boldsymbol{w_y} \in \mathbb{R}^{N_h \times 1}$ such that the correlation between the resulting linear combinations of canonical variates $\boldsymbol{x} = \boldsymbol{w_x}^T \boldsymbol{X}$ and $\boldsymbol{y} = \boldsymbol{w_y}^T \boldsymbol{Y}$ is maximized as:

$$
\begin{aligned}
\rho &= \arg \max_{\boldsymbol{w_x}, \boldsymbol{w_y}} \frac{E\left[\boldsymbol{x}\boldsymbol{y}^T\right]}{\sqrt{E\left[\boldsymbol{x}\boldsymbol{x}^T\right] E\left[\boldsymbol{y}\boldsymbol{y}^T\right]}} \\
&= \arg \max_{\boldsymbol{w_x}, \boldsymbol{w_y}} \frac{\boldsymbol{w_x}^T \boldsymbol{X}\boldsymbol{Y}^T \boldsymbol{w_y}}{\sqrt{\boldsymbol{w_x}^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w_x}} \sqrt{\boldsymbol{w_y}^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w_y}}}
\end{aligned}
\tag{1}
$$

Maximizing correction between input and reference signals (1) can be formulated and solved using generalized eigenvalue problem.

In baseline CCA based frequency detection method, the ideal sine-cosine reference signals, i.e., $\boldsymbol{Y}_i \in \mathbb{R}^{2N_h \times N_s}$ ($i = 1, 2, \ldots, N_f$), are created with a combination of sine-cosine signals of stimulus fundamental and different harmonics frequencies as follows [19]:

$$
\boldsymbol{Y_i} = \begin{pmatrix} \sin(2\pi f_i t) \\ \cos(2\pi f_i t) \\ \ldots \\ \sin(2\pi N_h f_i t) \\ \cos(2\pi N_h f_i t) \end{pmatrix}, \ t = \frac{1}{f_s}, \frac{2}{f_s}, \ldots, \frac{N_s}{f_s}
\tag{2}
$$

where f_s is the sampling rate. Then, the maximum correlation coefficient ρ_i can be computed between input EEG sample $\bar{\boldsymbol{X}} \in \mathbb{R}^{N_s \times N_s}$ and each reference $\boldsymbol{Y}_i$, $i = 1, 2, \ldots, N_f$, respectively where N_f is the number of stimulus frequencies. Finally, the SSVEP target frequency f can be identified as the frequency of the reference $\boldsymbol{Y}_i$ with the maximum correlation among all stimulus frequencies in Eq. (3):

$$
f_{\text{target}} = \max_f \rho(f), \ f = f_1, f_2, \ldots, f_{N_f}
\tag{3}
$$

2.2 DTW Temporal Alignment

DTW is a time-series alignment method that finds the non-linear temporal alignment mapping between input multi-dimensional input data $\boldsymbol{X} \in \mathbb{R}^{N_c \times N_s}$ and templates $\boldsymbol{Y} \in \mathbb{R}^{2N_h \times N_s}$ by minimizing the distance metric between them. DTW provides efficient time-series similarity measures regardless of shifting and distortion in time for alignment matching between the templates and target signals [26]. Comparison between two time-series uses local distance measure $d(x_k, y_k)$ to determine similarity if the distance metric is small (low cost) and dissimilarity if distance is large (high cost). For each pairs of input X and Y, the distance matrix $D \in \mathbb{R}^{N_s \times N_s}$ is computed to find the alignment between X and Y. The distance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be computed by minimizing path lengths P_x and P_y as

$$
\min_{P_x, P_y} d_{m,n} = \sum_{k=1}^{K} \|x_{k,m} - y_{k,n}\|
\tag{4}
$$

3

The warping path K is a sequence $p = [p_1, ..., p_K]$ with $p_k = (m_k, n_k) \in [1 : M] \times [1 : N]$ that satisfies three constraints to find the optimal solution in polynomial time as following.

- Boundary: $p_1 = [1, 1]$ *and* $p_K = [m_k, n_k]$.
- Monotonicity: $m_1 \leq ... \leq m_k$ *and* $n_1 \leq ... \leq n_k$.
- Continuity: $p_{k+1} - p_k \in ([0, 1], [1, 0], [1, 1])$
  for $k \in [1 : K - 1]$.

The optimal path length K that includes indexes to align the samples, is automatically selected by DTW algorithm having minimum distance among all possible warping paths. The $O(XY)$ time complexity of DTW can be reduced through multi-scaling, warping path constraints, etc to achieve linear computational efficiency. For SSVEP recognition task, DTW's stimulus templates can easily be constructed by consisting of SSR characteristics with narrow band responses to fundamental and harmonics of the stimulus frequencies. Unlike conventional time series matching problems, SSVEP classification only requires short data length of sub-second duration resulting in total numbers of samples less than or equal to 250 samples where $f_s = 250Hz$.

For m stimulus frequencies, each target can be represented by DTW as $d(f_m)$ and the target frequency, f_{target} is the frequency with minimum distance in comparison among reference templates as

$$f_{\text{target}} = \min_f d(f), \ f = f_1, f_2, \ldots, f_{N_f} \tag{5}$$
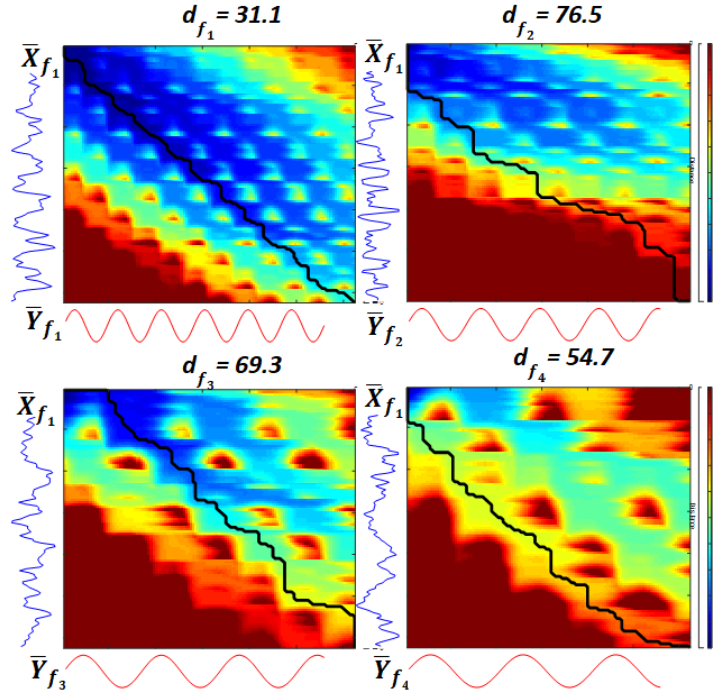

Figure 1: An example of SSVEP Target Detection with 4-class OpenBMI data-set by temporal alignment matching between input EEG and reference templates. For simplicity, only a single EEG channel with template of first harmonic was illustrated. The black-colored diagonal line shows distance measure of similarity between input and template signals where single-valued distance metric, d_{f_i} can be derived.

Figure. 1 illustrates computing distance d_i where $i = 1, 2, 3, 4$ for each stimulus frequency by comparing input EEG signals $\hat{X}_{f_1}$ with each template frequency, Y_{f_i}. As expected, d_1 is minimum distance compared with d_{n-1} resulting in correct frequency detection.

2.3   Combined Spatial-Filtering and Temporal Alignment

The proposed method, CSTA, consists of four steps in correctly classifying SSVEP responses among multiple stimuli as shown in Fig. 2 where EEG and reference signals at the i-th stimulus frequency f_i as X and $Y_i(i = 1, 2, \cdots, N_f)$,

respectively. As a first step, basic CCA is used to find maximum linear relationship between X and Y_i. The maximum correlation coefficients, ρ_i of respective stimulus frequencies, f_i are considered for spatial filtering based decision as in Eq. (3). Secondly, the canonical weight vector, w_x is used to transform input EEG signals X into canonically correlated EEG signals, i.e., $\hat{X} = X^T w_x$, as the input to DTW temporal alignment mapping step.
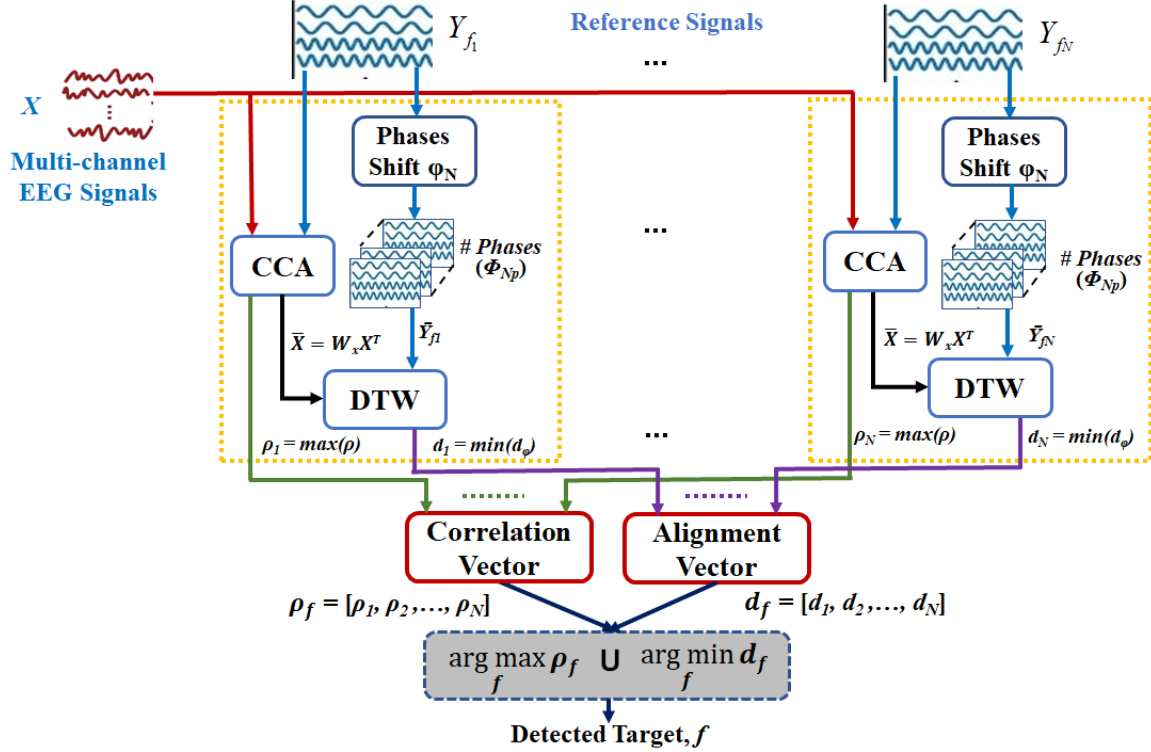

Figure 2: Diagram of the proposed method, CSTA in SSVEP Classification.

By concatenating ρ_i from each i-th stimulus frequency, correlation vector $f_\rho = [\rho_1, \rho_2, ..., \rho_{N_f}]$ was obtained from linear spatial filtering operation. Secondly, the CCA reference signal Y_i is expanded into multiple reference templates, $\hat{Y}_i^k$ by shifting phases specified as follows:

$$\phi_k = \phi_0 + \triangle\phi N_k, \; N_k = 0, 1, ..., k \tag{6}$$

where ϕ_0 is the phase used in stimulus with joint frequency and phase modulation coding [29] and zero if stimulus is only frequency modulation coding [30]. Multiple phases ϕ_k modified CCA reference signals Y allows to include k different phases as template signals, $\hat{Y}$ in DTW temporal alignment matching operation as follows:

$$\hat{Y}_i^k = \begin{pmatrix} \sin(2\pi f_i t) + \phi_k \\ \cos(2\pi f_i t) + \phi_k \\ \cdots \\ \sin(2\pi N_h f_i t) + \phi_k \\ \cos(2\pi N_h f_i t) + \phi_k \end{pmatrix} \tag{7}$$

Then, temporal similarity between $\hat{X}$ and $\hat{Y}_i^k$ is determined by minimizing the warping path length using DTW for each frequency and phased-template as $d(\phi_k, N_f)$ according to Eq (5). For comparison among distance metrics, non-normalized distances, $\hat{d}(\phi_k, i)$ computed from temporal alignment using multiple references are normalized resulting $\hat{d}$. For each i-th stimulus frequency, the minimum distance d_i is determined from multi-phased normalized warped distances $\hat{d}(\phi_k, i)$ according to k phase shifts.

$$d_i = \min_k \hat{d}(\phi_k, i) \tag{8}$$

Similarly, the distance vector f_d is formed by concatenating normalized minimum distances of all stimulus frequencies $[d_1, d_2, ..., d_{N_f}]$. The final target classification uses intermediate decision vectors from CCA, f_ρ and DTW, f_d to

determine the detected target, f_{CSTA}. We used complementary fusion to combine decision outcomes from maximization of CCA spatial filtering and minimization of DTW temporal alignment as shown in Eq. (9). Further, the correct class detection is governed by two decision rules: (1) matching with ground-truth frequency f_c with detected label(s) from either or both linear correlation, f_ρ and non-linear alignment, f_d operations; (2) at least twenty-five percents of frequency detection agreement between both operations. But fusion rule can easily be varied from this simple fusion logic to weighted voting fusion between d_f and ρ_f depending on the performance criteria and data-specific requirements.

$$f_{CSTA} = \arg\max_f \hat{\rho}_f \cup \arg\min_f \hat{d}_f \tag{9}$$

3   Experiment Design and Evaluation Analysis

For evaluating the performance of the proposed training-free method, we select two SSVEP data-sets to compare with existing SOTA methods including one training-free method, FBCCA [20] and two training-based methods, TRCA [13] and IT-CCA [21]. We used similar data and parameters settings with the same data analysis pipeline in offline evaluation analysis. We will explain briefly about two SSVEP data-sets and three SOTA methods in below sub-sections.

3.1   SSVEP Data-sets

The first SSVEP data-set, OpenBMI, was collected by Korea University [30]. The second data-set was an open access benchmark data-set, HS-SSVEP, by Tsinghua University [29]. These data-sets have distinct stimulus characteristics and experiment conditions. OpenBMI data-set, 4-class SSVEP, has 10 times lower spatial resolution than HS-SSVEP data-set, 40-class SSVEP. The stimuli in OpenBMI are generated using the number of frames in a period according to monitor screen's refresh rate such as 8.57Hz (7 frames per period of 60Hz), whereas those of HS-SSVEP are generated using frequency approximation approach [31].

OpenBMI data-set is a representative of control task where four targets SSVEP interface is coded with four distinct frequencies, such as, 5.25 Hz, 6.67 Hz, 8.57 Hz, 12 Hz [30]. The EEG data were recorded using the 62 channels EEG at 1000 Hz sampling rate. Fifty-four subjects participated in the experiment with a total of four sessions. Each session has 100 trials corresponding to four stimuli that were presented in a random order. So each stimulus has 25 trials per session. In each trial, the stimulus duration was 4 s, which was followed by 6 s rest interval. To match with the second data-set in SSVEP analysis, all data were down-sampled to 250 Hz.

HS-SSVEP data-set is a representative of visual spelling task [29]. Its interface consists of forty targets by coding unique frequency and phase information for each stimulus. The stimuli are coded with frequencies of 8-15.8 Hz with an interval of 0.2 Hz in combination with four distinct phases of $0, 0.5\pi, \pi$ and $1.5pi$ radians [29]. The EEG data from thirty-five healthy subjects were recorded from 64 channels at 1000 Hz sampling rate. For each subject, six blocks of data in single session were recorded. In each block, forty trials corresponding to each stimulus frequency were conducted in a random order. So each stimulus frequency has six trials. The duration of each trial is 6 s, in which the first and last 0.5 s were used for visual cue and rest interval respectively. The provided EEG data for each trial are downsampled at 250 Hz.

In our analysis, the time window or response time is set from 0.2 to 1 s with an increment of 0.1s. We only consider data from nine channels at occipito-parietal area: namely, PO8, Pz, PO7, PO4, POz, PO3, O2, Oz, O1 available in both data-sets. All data epochs are preprocessed through zero-phase IIR band-pass filter from 4Hz to 75Hz, and from 7Hz to 90Hz for OpenBMI and HS-SSVEP data-set respectively. This evaluation allows to assess the performance of different methods with distinct SSVEP characteristics such as the number of targets, stimulus coding, frequency range, etc. The number of phase, k, in CSTA methods is set empirically at 8 and 17 in OpenBMI and HS-SSVEP data-set respectively.

3.2   Training-free Method

Filter-Bank CCA (FB-CCA) method is still a calibration-free method as no training is required for each subject. In this method, multi-channel EEG signals are preprocessed with pre-defined band-pass filters to extract sub-band components [20]. By applying pre-constructed equal bandwidth filter banks, EEG signals X are decomposed into 'n' sub-bands resulting in X_{SB}^n components. Standard CCA is applied to each sub-band comparing with ideal sine-cosine reference signals Y to compute correlation vector ρ_f^n. The target frequency, f_{target} is the weighted sum of the square of the correlation coefficients from all sub-band N components.

$$f_{target} = \sum_{n=1}^{N} w(n).(\rho_f^n)^2. \tag{10}$$

where $w(n) = n^{(-a)} + b$. In our analysis, we used n sub-bands where each sub-band starts with unique m frequencies but ends with about five or six times of f_{max} of stimulus frequencies of the target data-set. The constants a and b from weight vector w_n are empirically identified through parameters grid search for each data-set using formula defined in [20]. The starting frequency is set at $n \times 5$ Hz for n^{th} sub-bands with same ending frequency at 72Hz for OpenBMI data-set. For HS-SSVEP data-set, the starting frequency for n^{th} sub-bands is set at $n \times 8$ Hz with same ending frequency at 88Hz [20]. The a and b coefficient pairs of weight vector w_n are set at [1.25, 0.25] and [3.0, 1.0] for HS-SSVEP and OpenBMI data-set respectively after performing parameters grid search [20, 32]. The number of FB (n) is set at 5 for both data-sets [14].

3.3 Training-based Methods

Individual Template-CCA (IT-CCA) is training-based method as calibration data are required to create subject-specific references [21]. Subject-specific templates are created by averaging EEG signals across multiple trials from each subject [33]. This method achieved higher performance in comparison with various improved CCA methods using 12-class SSVEP data-set [21]. Instead of single maximum correlation value among ρ in Eq. (1), each stimulus frequency is represented by the correlation coefficients between the linear combination of test EEG samples $\bar{X} \in \mathbb{R}^{N_c \times N_s}$, subject-independent reference signals $Y_i \in \mathbb{R}^{2N_h \times N_s}$ and individual subject-specific templates $Z_i \in \mathbb{R}^{N_c \times N_s}$ ($i = 1, 2, \ldots, N_f$) using different spatial filters. For the i-th frequency, the correlation coefficient ρ_i was derived as the summation of correlation vectors $r_i(m)$ as follows [21].

$$\rho_i = \sum_{m=1}^{5} \text{sign}(r_i(m)) \cdot (r_i(m))^2 \qquad (11)$$

where $\text{sign}(\cdot)$ was used to retain discriminative information, positive or negative correlation, from correlation vectors r_i. The target frequency of input EEG $\bar{X}$ was then identified by selecting the frequency with maximum correlation values similar to Eq. (3).

Task Related Component Analysis (TRCA) is another subject-specific training-based method that exploits reliable frequency response (SSR) characteristics of SSVEP with respect to stimulus frequency. Nakanishi et al proposed TRCA spatial filtering by exploiting task-related components by linear and weighted sum of multiple time courses that optimizes the maximum covariance among trials [13]. TRCA essentially learns spatial filters to extract task-related components by maximizing the reproducibility of SSR from multiple trials [34]. So TRCA optimizes the weight vector $w \in \mathbb{R}^{N_c \times 1}$ from the assumption that temporal profiles of task-related components $y^{(1)}, y^{(2)}, \cdots, y^{(N_h)}$ exhibits a maximal temporal similarity among trials. With N_f spatial filters, $w_n \in \mathbb{R}^{N_c \times 1}$ and the template signals $Y_n \in \mathbb{R}^{N_c \times N_s}$, where $n = 1, \cdots, N_f$, we can quantify the linear correlation λ_n between the each template signal and a test sample $X \in \mathbb{R}^{N_c \times N_s}$ using one-dimensional correlation analysis. Finally, the frequency of the test sample X could be determined from resultant correlation coefficient λ_n as:

$$f_{target} = \max_n \left(\rho(w_n^T X, w_n^T Y_n) \right), \ n = 1, 2, \cdots, N_f \qquad (12)$$

We used leave-one-trial-out cross validation in our data-sets for both training-based methods [13]. Out of N_t trials, $N_t - 1$ trials are used to train individual templates, Z_i, i.e. 24 trials and 5 trials respectively for OpenBMI and HS-SSVEP data-set, and remaining 1 trial repetitively to compute the mean accuracy for performance comparison. The FB extension shows performance improvements with other baseline methods including CCA in evaluation study with HS-SSVEP data-set [32]. FB as preprocessing step improved the performance of both calibration-free and calibration-based methods in previous studies [13, 14, 32]. So we also apply FB preprocessing in SOTA methods in order to compare the performance of the CSTA method with similar FB extension.

4 Results and Discussion

We compared the performance between CSTA, and 3 SOTA methods, i.e., FBCCA, IT-CCA and TRCA using Accuracy and ITR metrics [35, 13]. The number of harmonics was set to five in all methods to include the fundamental and multiple harmonic components of SSVEP responses [15]. For ITR computation, we include cue interval of 2 s and 0.5 s time window for OpenBMI and HS-SSVEP data-set respectively [9]. TRCA is commonly used as benchmark evaluation method in comparison with newly proposed training-based methods using HS-SSVEP data-set [13, 29]. FB preprocessing step applied to TRCA, IT-CCA and CSTA methods are named as FB-TRCA, FB-ITCCA and FB-CSTA respectively. This FB extension allows to evaluate how FB extension improves performance and behaves differently with those methods in two data-sets.

For statistical evaluation analysis, we used one-way ANOVA with repeated measures analysis with Greenhouse-Geisser spherical correction after testing sphericity assumption using Mauchly's test of sphericity for performance comparison among multiple methods [20]. All post-hoc pairwise comparisons between different method pairs were computed with Bonferroni correction. Two alpha levels of 0.001 and 0.05 are used for statistical significance testing. Also, two-sided non-parametric Wilcoxon rank sum test was used to test significant difference between methods with option of with or without FB extension.
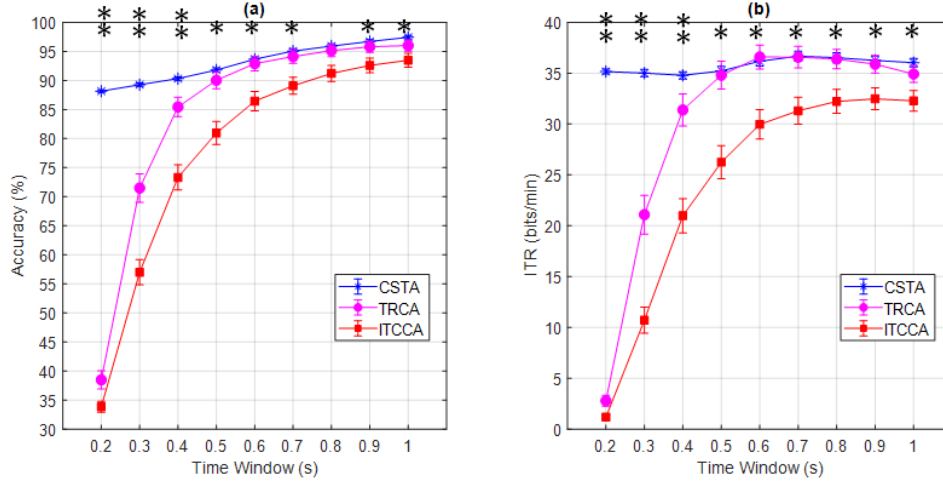

Figure 3: Performance comparison of CSTA with two training methods (TRCA, IT-CCA) for OpenBMI data-set (a) Accuracy (b) ITR using one-way ANOVA with repeated measures after spherical correction. At specific time window, ** indicates significant difference among methods at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise no statistical significant difference among methods. Error bar indicates standard error.

From one-way ANOVA with repeated measures with OpenBMI data-set, the results show high statistically significance ($p < 0.001$) in mean accuracy and ITR among methods from 0.2 s to 0.4 s time windows. The significant in mean accuracy and ITR was achieved ($p < 0.05$) from 0.5 s to 1.0 s except not statistical significant difference in accuracy at 0.8 s ($F(2, 104) = 3.01, p = 0.052$). Figure. 3 shows that CSTA performs either better or similar accuracy and ITR in comparison with two training-based methods in time windows from 0.2 s to 1.0 s. Interestingly, CSTA results show less variability in mean accuracy among 0.2 s to 1 s time windows. Averaging accuracy per subject across time windows highlights superior performance of CSTA with 95.15 ± 3.4 % overall accuracy compared with 84.38 ± 18.89 % and 77.55 ± 20.13 % of TRCA and IT-CCA respectively. The mean accuracy difference between CSTA and other methods at 1 s showed $1 - 4\%$ on where maximum accuracy was achieved for each method. But the mean accuracy difference of CSTA is about $49 - 54\%$ higher than other methods at 0.2 s time window. This highlights better performance improvements in CSTA can be seen at short time windows. Similar to accuracy, highly significant mean ITR among methods ($p < 0.001$) was achieved from 0.2 s to 0.4 s time window. The significant in mean ITR ($p < 0.05$) among methods was shown from 0.5 s to 1 s. The maximum mean ITR of 36.67 bpm is achieved at 0.7 s time window with CSTA method. But TRCA achieved maximum mean ITR of 36.59 bpm at 0.6 s whereas that of IT-CCA is 32.48 bpm at 0.9 s. The low mean ITR results in OpenBMI data-set are mainly due to inclusion of long cue duration of 2 s used in stimulus presentation [30].

Table 1: Mean standard error comparison ($\mu \pm \sigma$) over time windows of three methods using two-data-set

Data-set	Metric	CSTA	TRCA	IT-CCA
OpenBMI	ACC	0.3 ± 0.1	1.4 ± 0.5	1.6 ± 0.4
	ITR	0.41 ± 0.08	1.14 ± 0.42	1.2 ± 0.44
HS-SSVEP	ACC	2.11 ± 0.5	3.56 ± 0.5	3.37 ± 0.5
	ITR	7.65 ± 1.56	11.75 ± 2.32	10.91 ± 2.1

Table. 1 presents mean standard errors of accuracy and ITR for comparison among methods across all time windows derived in both data-sets as shown in Figure. 3 and Figure. 4. Apparently, CSTA method has lower standard errors than TRCA and IT-CCA methods;resulting less variability in mean performance across time window in both data-sets as shown in Table. 1. The possible reason for consistent performance across time window in CSTA might be due to low

Table 2: Post hoc pairwise comparison of Accuracy among method pairs without FB. At specific time window, ** indicates significant difference at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise p-value for no statistical significant difference.

Methods	data-sets	0.2 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s	0.8 s	0.9 s	1 s
CSTA Vs TRCA	OpenBMI	**	**	*	0.599	1.0	0.994	1.0	0.689	0.186
	HS-SSVEP	**	**	**	**	*	0.87	1.0	1.0	1.0
CSTA Vs ITCCA	OpenBMI	**	**	**	**	**	**	**	**	**
	HS-SSVEP	**	**	**	**	**	0.116	0.493	1.0	1.0
TRCA Vs ITCCA	OpenBMI	**	**	**	**	**	**	**	**	**
	HS-SSVEP	*	*	0.55	1.0	0.211	0.295	0.366	*	0.099

stimulus spatial resolution on screen (only 4-class) with frequency only coded stimuli and high adjacent inter-stimulus frequency difference (2.18 ± 1.13 Hz). This neighboring frequency difference is more than ten times higher than that of HS-SSVEP data-set, 0.2 Hz. With distinct and well-separate stimulus frequencies, combination of linear correlation with non-linear temporal alignment matching recognizes target frequencies more accurately. Although phases from steady-state responses are stable, the actual phase information will possibly deviate from stimulus frequencies due to visual latency from retina to brain pathway [36]. As both CSTA and IT-CCA methods include linear correlation measures in frequency recognition, the extension of multi-phased templates with temporal alignment in CSTA contributed to above superior performance outcomes. According to post-hoc pairwise mean accuracy comparison results as shown in Table 2, CSTA performs significantly better ($p < 0.001$) than IT-CCA in all time windows. Similarly, TRCA has significantly better accuracy than IT-CCA ($p < 0.001$) in all time windows. But CSTA only performs significant better than TRCA from 0.2 s to 0.4 s and no significant difference between them from 0.5 s to 1 s as shown in Table 2. Similar significant results as accuracy, except at 0.4 s time window ($p = 0.059$), achieved in pairwise comparison between pairs of three methods for ITR as shown in Table 3.
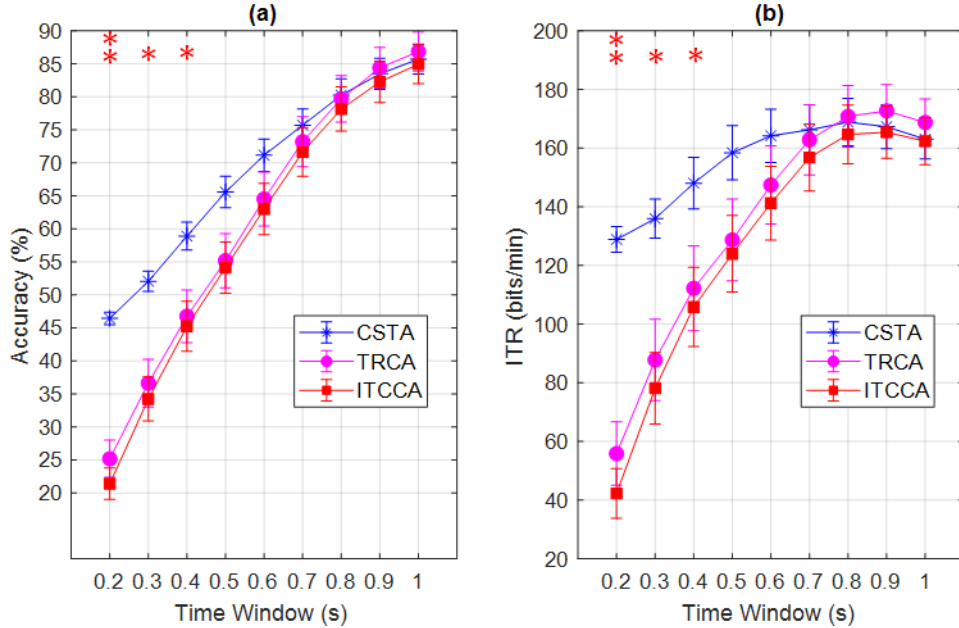

Figure 4: Mean performance comparison of CSTA with two training methods for HS-SSVEP data-set (a) Accuracy (b) ITR using one-way ANOVA with repeated measures with spherical correction. At specific time window, ** indicates significant difference among methods at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise no statistical significant difference. Error bar indicates standard error.

From applying one-way ANOVA with repeated measures in HS-SSVEP data-set, different statistically significant results in mean accuracy among methods in short time windows as show in Fig.4. Only high significant difference at 0.2 s ($p < 0.001$) and significant difference ($p < 0.05$) from 0.3 s-0.4 s among methods were achieved. Similar statistical significant results of mean ITR among methods similar to mean accuracy can be seen in Figure.4. Unlike OpenBMI

Table 3: Post hoc pairwise comparison of ITR among method pairs without FB. At specific time window, ** indicates significant difference at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise p-value for no statistical significant difference result.

Methods	data-sets	0.2 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s	0.8 s	0.9 s	1 s
CSTA Vs TRCA	OpenBMI	**	**	0.059	1.0	1.0	1.0	1.0	1.0	0.193
	HS-SSVEP	**	**	**	*	0.098	1.0	1.0	0.827	0.63
CSTA Vs ITCCA	OpenBMI	**	**	**	**	**	**	**	**	**
	HS-SSVEP	**	**	**	**	**	0.252	1.0	1.0	1.0
TRCA Vs ITCCA	OpenBMI	**	**	**	**	**	**	**	**	**
	HS-SSVEP	*	*	0.276	0.662	0.295	0.137	0.107	*	*

data-set, no significant difference among methods from 0.5 to 1.0 s in mean accuracy and ITR performance. Pairwise ad-hoc comparison of accuracy and ITR between methods of HS-SSVEP data-set in Table 2 showed differently in comparison with OpenBMI data-set as shown in Table 3. Different levels of statistical significant results can be seen in both mean accuracy (Only significant difference from 0.2 s-0.6 s) and ITR (Only significant difference from 0.2 s-0.5 s) comparison between CSTA and TRCA methods.But highly significant results ($p < 0.001$) between CSTA and IT-CCA in both accuracy and ITR in 0.2 s to 0.6 s. These statistical results show variability in significant performance difference among methods in HS-SSVEP data-set compared with OpenBMI. But CSTA method consistently performs either better or comparable with both training methods in short time windows. Similar to openBMI dataset, mean performance of CSTA method has less variability compared with other methods across all time windows as shown in Table. 1. This highlights the contribution of temporal alignment matching with multi-phased templates and complementary fusion between correlation and alignment decision vectors. Comparison of all three methods above did not include FB as pre-processing step.

Generally, FB extensions of different methods in SSVEP detection showed $10 - 30\%$ accuracy improvements with HS-SSVEP data-set from 0.25 s to 5 s time window [32]. But the most notable performance improvement was observed in longer time window after 1 s segment with FB extension [32]. Although FB extension does not require user-specific calibration step, it still requires to perform a grid search for finding optimal weight coefficients in Eq. (10) for each method in each data-set [32]. This is one limitation of FB extension to optimize weight vector for each data-set for optimal performance improvements.

FBCCA method shows better accuracy improvements in longer time window of more than 0.5 s compared with baseline CCA in both data-sets. But its accuracy is close to baseline accuracy of 25% and 2.5% in 0.2 - 0.3 s time windows in both data-sets as shown in Fig. 5. With one-way ANOVA with repeated measures analysis, highly significant difference ($p < 0.001$) among these methods can be seen from 0.2 s to 1.0 s time windows. This shows exploiting phase information with non-linear temporal alignment achieve high performance in sub-second response time. As FB operation exploits the existence of multiple harmonics from SSR in frequency recognition, longer time window is required for improving performance as can be seen in Fig. 5.

One-way ANOVA with repeated measures results show less significant performance difference among methods with FB extension in Fig. 6 compared with non-FB extensions using both data-sets in Figs. 3 and 4. The significant difference among FB extended methods can only be found in less than 0.7 s time window, i.e. 0.2 s-0.3 s ($p < 0.001$) and 0.4 s-0.6 s ($p < 0.05$) respectively, in OpenBMI data-set 6. But only significant difference ($p < 0.05$) among methods can be seen at 0.2 s in HS-SSVEP data-set. Table 4 shows mean accuracy comparison of FB extended methods of both data-sets from 0.2 s to 1.0 s time window. But overall accuracy improvements can be seen in all methods at different levels in both data-sets. The mean accuracy improvements in OpenBMI data-set show slight differences of 0.44 ± 0.39 %, 1.12 ± 2.64 % and 2.1 ± 0.8 % for CSTA, TRCA an IT-CCA methods respectively. In HS-SSVEP data-set, mean accuracy improvements of 4.76 ± 1.48 %, 10.36 ± 3.87 % and 9.5 ± 3.32 % respectively for above three methods. The proposed method CSTA gains the least accuracy improvements by FB extension in both data-sets. FB improvements can be clearly seen in HS-SSVEP data-set compared with OpenBMI data-set especially in TRCA and IT-CCA methods. The possible reason for higher performance improvement is due to equal inter-stimulus frequency difference of 0.2 Hz in HS-SSVEP data-set compared with unequal inter-stimulus frequency difference in OpenBMI data-set. Pairwise comparison of CSTA, TRCA and ITCCA methods with and without FB extensions at different time windows validates our assertions as shown in Table. 5.

We also use R^2 coefficients to compare the features discrimination between target and non-target stimuli of different methods [13]. The high R^2 value means good features discrimination, better features representations between targets
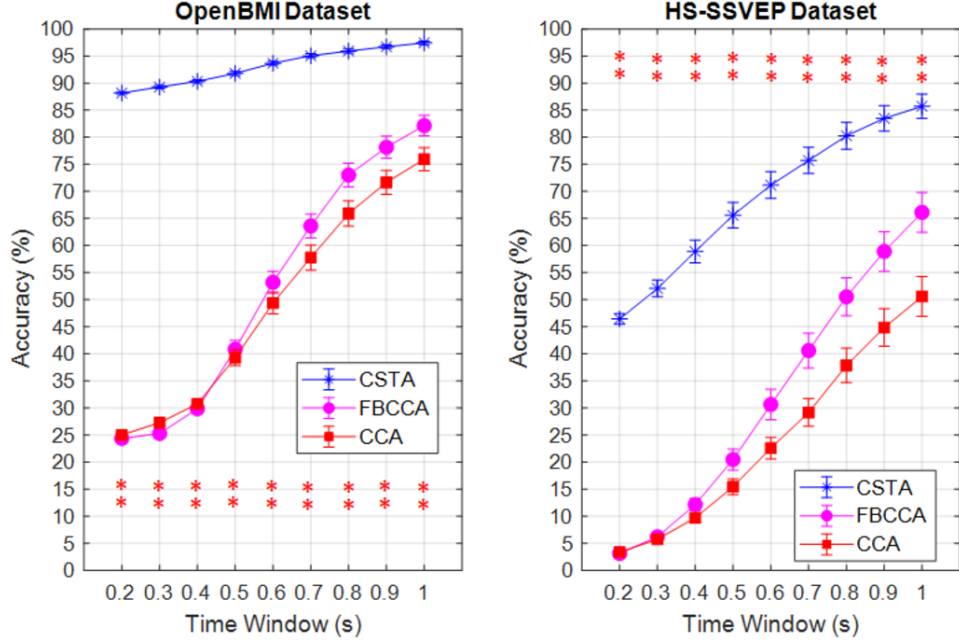
Figure 5: Mean Accuracy comparison among CSTA with training-free CCA and FB-CCA methods at different time windows using one-way ANOVA with repeated measures with spherical correction. Left and right graph shows results from OpenBMI and HS-SSVEP data-set respectively. Here, ** indicates significant difference among methods at $p < 0.001$ at specific time window. Error bar indicates standard error.

and non-targets, of the method. The results in Figure. 7 show CSTA has better features discriminative capability but has less features variability than other methods in both data-sets in each frequency.

This can be further validated by visualizing multi-phased non-linear temporal alignment features used in CSTA method in different time windows. We apply t-Distributed Stochastic Neighbor Embedding (t-SNE) to transform and visualize
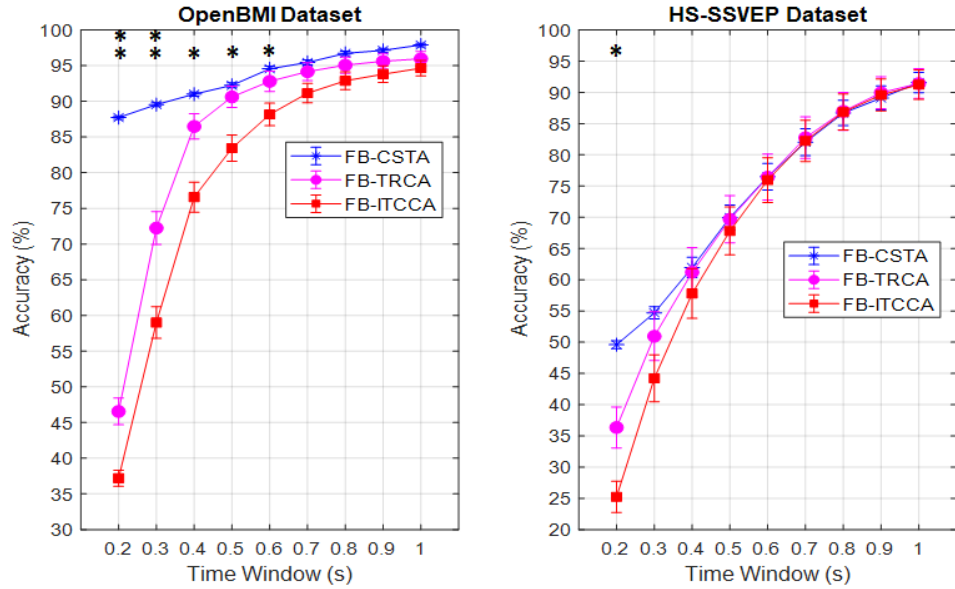

Figure 6: Accuracy comparison among three methods with FB extensions in OpenBMI (Left graph) and HS-SSVEP (Right graph). At specific time window, ** indicates significant difference at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise ' ' for no statistical significant difference. Error bar indicates standard error.

11

Table 4: Post hoc pairwise comparison of accuracy among methods with filter-bank extension. At specific time window, ** indicates significant difference at $p < 0.001$, * indicates significant difference at $p < 0.05$, otherwise p-value for no statistical significant difference result.

Methods	data-sets	0.2 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s	0.8 s	0.9 s	1 s
CSTA Vs TRCA	OpenBMI	**	**	*	0.774	0.513	0.745	0.284	0.28	0.114
	HS-SSVEP	**	0.714	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CSTA Vs ITCCA	OpenBMI	**	**	**	**	**	*	**	*	*
	HS-SSVEP	**	*	0.359	0.971	1.0	1.0	1.0	1.0	1.0
TRCA Vs ITCCA	OpenBMI	**	**	**	**	**	**	**	*	*
	HS-SSVEP	**	**	*	0.123	1.0	1.0	1.0	1.0	1.0

Table 5: Two-sided non-parametric Wilcoxon rank sum tests of comparison between with and without FB pre-processing of three methods. At specific time window, ** indicates significant difference at $p < 0.001$, * indicates significant difference $p < 0.05$, otherwise shows p-value if no statistically significant difference.

Methods	data-sets	0.2 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s	0.8 s	0.9 s	1 s
CSTA	OpenBMI	0.177	0.463	0.107	0.493	0.104	0.491	0.109	0.407	*
	HS-SSVEP	**	*	0.077	0.139	0.085	0.061	*	0.066	*
TRCA	OpenBMI	**	0.89	0.561	0.612	0.873	0.929	0.961	0.851	0.813
	HS-SSVEP	*	*	*	*	*	*	0.051	0.066	0.073
ITCCA	OpenBMI	*	0.505	0.252	0.333	0.386	0.217	0.233	0.352	0.33
	HS-SSVEP	0.14	*	*	*	*	*	*	*	*

multi-dimensional features from both methods into 2-dimensional feature presentations [37]. Figure. 8 proves that the features identified by CSTA method has similar separability among different frequencies between 0.3 s and 0.8 s time windows although slightly better inter-frequency separation at 0.8 s. We believe this good inter-class separability enables better performance in short time windows compared with SOTA methods in OpenBMI and HS-SSVEP as shown in Fig. 3 and Fig. 4 respectively.

The performance of TRCA can be further improved using ensembled and hierarchical extensions though such extensions are not used in our comparison [13, 15]. Another training-based approach constructs spatiotemporal beamformers from calibration trials, features that optimizes weights per stimulus, by minimizing the beamforming variance [38]. But, the
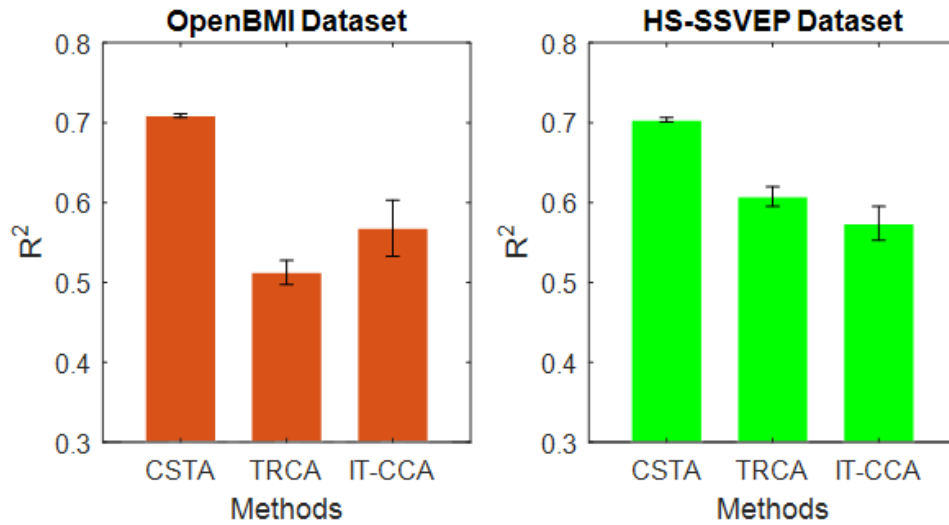

Figure 7: Features comparison among methods with mean R-squared values for SSVEPs at 12 Hz and 8.4 Hz for OpenBMI and HS-SSVEP data-set respectively. Time window is set at 0.3 s. Error bar indicates standard errors.
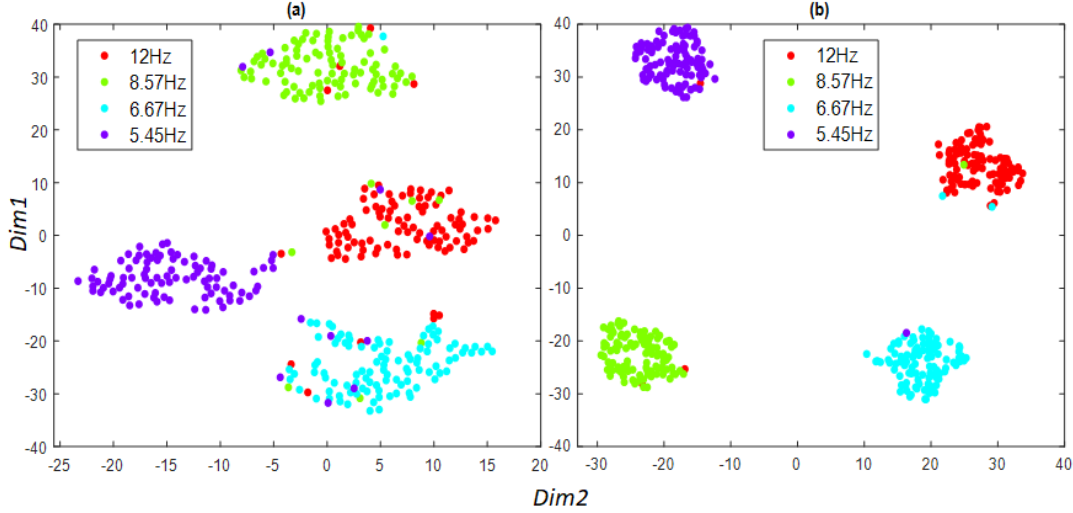
Figure 8: Clusters-based visualization of features, after dimensional reduction using t-SNE to two dimensions for simple illustration, derived from CSTA method for 4-class openBMI data-set at (a) 0.3 s (b) 0.8 s time window.

performance of these training methods are evaluated in subject-specific leave-one-trial-out cross validation. In order to evaluate in a subject-independent manner, there is still no empirical evaluation of these training-based methods similar to training-free methods using leave-one-subject-out scenario. In this manner, overall performance evaluation is similar and comparable to that of subject-independent SSVEP recognition [20, 27]. Moreover, most training-based methods such as TRCA, ITCCA exploits spatial filters that optimize linear weights of SSVEP responses assuming that consistent data characteristics across trials, sessions and days for each subject. With underlying sources of non-linear SSVEP responses and non-linearity of scalp EEG nature, linear spatial filtering with optimal weights methods might possibly miss-classify some SSVEP responses [23]. By combining linear spatial filtering with non-linear temporal alignments, CSTA method can offer high detection and consistent performance in short time window.

The only limitation of CSTA method is to empirically specify the numbers of phases, k, used in multi-phased templates for temporal alignment. This will raise a question of how to systematically optimize k parameter in CSTA method as well as how different k values affect the SSVEP classification performance. Due to relatively high accuracy achieved with conventional data-driven approaches, deep learning approaches are not commonly used compared with other BCI modalities like motor imagery [39]. Recently, deep learning based SSVEP recognition using custom 12 stimuli showed significant performance improvements compared with CCA methods in 1 s time window [40]. We intend to further evaluate performance of CSTA methods with subject-independent cross validation of SOTA training-methods including deep learning in short time windows. Furthermore, we will continue to improve CSTA performance through subject-specific templates like IT-CCA [21] and improved time-series alignment mapping methods [41]. Notably, extension of TRCA method by exploiting multiple stimuli's characteristics show further performance improvements compared with baseline and ensembled TRCA [42]. Nevertheless, our work highlights the subject-independent training-free method can still perform better or similar compared with training-based methods for SSVEP frequency recognition in short response time.

5 Conclusion

For achieving a fast SSVEP-based BCI, frequency recognition methods must provide high accuracy with less performance variability in sub-second response time. It is highly desirable that such methods require no subject-specific training to reduce calibration time ensuring high usability. Existing methods only improve accuracy by either exploiting subject-specific calibration in short response time, training-based methods, or relying on long time window, training-free methods. In the current study, we propose a CSTA method leveraging linear correlation and non-linear temporal alignment mapping with decision fusion to improve accuracy in short response time. The offline experimental evaluation shows significant performance improvements with CSTA compared with both training and training-free methods, especially in response time of less than 0.5 s. We validated that performance improvements on two SSVEP data-sets with different stimulus characteristics. We are hoping that the proposed CSTA method brings towards us

the realization of fast SSVEP-based BCI by fulfilling end-user needs of training-free approach for reliable frequency recognition in sub-second response time.

Acknowledgment

References

[1] Clemens Brunner, Niels Birbaumer, Benjamin Blankertz, Christoph Guger, Andrea Kübler, Donatella Mattia, José del R Millán, Felip Miralles, Anton Nijholt, Eloy Opisso, et al. BNCI horizon 2020: towards a roadmap for the BCI community. *Brain-computer interfaces*, 2(1):1–10, 2015.

[2] Vikram Shenoy Handiru, AP Vinod, and Cuntai Guan. EEG source imaging of movement decoding: The state of the art and future directions. *IEEE Systems, Man, and Cybernetics Magazine*, 4(2):14–23, 2018.

[3] Bin He, B. Baxter, B.J. Edelman, C.C. Cline, and W.W. Ye. Noninvasive brain-computer interfaces based on sensorimotor rhythms. *Proc. IEEE*, 103(6):907–925, 2015.

[4] Yosuke Kimura, Toshihisa Tanaka, Hiroshi Higashi, and Naoki Morikawa. SSVEP-based brain-computer interfaces using FSK-modulated visual stimuli. *IEEE Trans. Biomed. Eng.*, 60(10):2831–2838, 2013.

[5] Nannan Zhang, Yadong Liu, Erwei Yin, Baosong Deng, Lu Cao, Jun Jiang, Zongtan Zhou, and Dewen Hu. Retinotopic and topographic analyses with gaze restriction for steady-state visual evoked potentials. *Scientific reports*, 9(1):4472, 2019.

[6] Jing Jin, Brendan Z Allison, Eric W Sellers, Clemens Brunner, Petar Horki, Xingyu Wang, and Christa Neuper. An adaptive P300-based control system. *J. Neural Eng.*, 8(3):036006, 2011.

[7] Zhenghui Gu, Zhubing Chen, Jintao Zhang, Xichun Zhang, and Zhu Liang Yu. An online interactive paradigm for P300 brain-computer interface speller. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 27(2):152–161, 2019.

[8] Alexander M Dreyer, Christoph S Herrmann, and Jochem W Rieger. Tradeoff between user experience and BCI classification accuracy with frequency modulated steady-state visual evoked potentials. *Frontiers in human neuroscience*, 11:391, 2017.

[9] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-speed spelling with a noninvasive brain-computer interface. *Proc.Natl. Acad. Sci. USA.*, 112(44):E6058–E6067, 2015.

[10] Anthony M. Norcia, L. Gregory Appelbaum, Justin M. Ales, Benoit R. Cottereau, and Bruno Rossion. The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6):4–4, 05 2015.

[11] Y S Zhang, Peng Xu, Kaiwen Cheng, and Dezhong Yao. Multivariate synchronization index for frequency recognition of SSVEP-based brain-computer interface. *J. Neurosci. Meth.*, 221:32–40, 2014.

[12] Y S Zhang, Daqing Guo, Peng Xu, Yu Zhang, and Dezhong Yao. Robust frequency recognition for SSVEP-based BCI with temporally local multivariate synchronization index. *Cognitive Neurodynamics*, 10(6):505–511, 2016.

[13] M. Nakanishi, Y. Wang, X. Chen, Y. T. Wang, X. Gao, and T. P. Jung. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.*, 65(1):104–112, Jan 2018.

[14] Y S Zhang, Erwei Yin, Fali Li, Yu Zhang, Toshihisa Tanaka, Qibin Zhao, Yan Cui, Peng Xu, Dezhong Yao, and Daqing Guo. Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 26(7):1314–1323, 2018.

[15] Y S Zhang, Erwei Yin, Fali Li, Yu Zhang, Daqing Guo, Dezhong Yao, and Peng Xu. Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs. *Neural Networks*, 119:1–9, 2019.

[16] Noémie Hébert-Lalonde, Lionel Carmant, Dima Safi, Marie-Sylvie Roy, Maryse Lassonde, and Dave Saint-Amour. A frequency-tagging electrophysiological method to identify central and peripheral visual field deficits. *Documenta Ophthalmologica*, 129(1):17–26, 2014.

[17] Masaki Nakanishi, Yu-Te Wang, Tzyy-Ping Jung, John K Zao, Yu-Yi Chien, Alberto Diniz-Filho, Fabio B Daga, Yuan-Pin Lin, Yijun Wang, and Felipe A Medeiros. Detecting glaucoma with a portable brain-computer interface for objective assessment of visual function loss. *JAMA ophthalmology*, 135(6):550–557, 2017.

[18] Vaegan, Anmar M. A. Rahman, and Gordon F. Sanderson. Glaucoma affects steady state vep contrast thresholds before psychophysics. *Optometry and Vision Science*, 85(7), 2008.

[19] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.*, 54(6):1172–1176, 2007.

[20] Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. *J. Neural Eng.*, 12(4):046008, 2015.

[21] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLoS ONE*, 10(10):e0140703, 2015.

[22] Peng Yuan, Xiaogang Chen, Yijun Wang, Xiaorong Gao, and Shangkai Gao. Enhancing performances of SSVEP-based brain-computer interfaces via exploiting inter-subject information. *J. Neural Eng.*, 12(4):046006, 2015.

[23] Maciej Labecki, Rafal Kus, Alicja Brzozowska, Tadeusz Stacewicz, Basabdatta S Bhattacharya, and Piotr Suffczynski. Nonlinear origin of ssvep spectra—a combined experimental and modeling study. *Frontiers in computational neuroscience*, 10:129, 2016.

[24] Rosanne Zerafa, Tracey Camilleri, Owen Falzon, and Kenneth P Camilleri. To train or not to train? a survey on training of feature extraction methods for SSVEP-based BCIs. *J. Neural Eng.*, 15(5):051001, 2018.

[25] Meinard Müller. *"Dynamic Time Warping"*, pages 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[26] Skyler Seto, Wenyu Zhang, and Yichen Zhou. Multivariate time series classification using dynamic time warping template selection for human activity recognition. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 1399–1406. IEEE, 2015.

[27] Aung Aung Phyo Wai, Min-Ho Lee, Seong-Whan Lee, and Cuntai Guan. Improving the performance of SSVEP BCI with short response time by temporal alignments enhanced CCA. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 155–158. IEEE, 2019.

[28] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[29] Y. Wang, X. Chen, X. Gao, and S. Gao. A benchmark dataset for SSVEP-based brain-computer interfaces. *IEEE Trans.Neural Syst. Rehabil.Eng.*, 25(10):1746–1752, Oct 2017.

[30] Min-Ho Lee, O-Yeon Kwon, Yong-Jeong Kim, Hong-Kyung Kim, Young-Eun Lee, John Williamson, Siamac Fazli, and Seong-Whan Lee. Eeg dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience*, 8(5):giz002, 2019.

[31] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, Yasue Mitsukura, and Tzyy-Ping Jung. Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate. *PLOS ONE*, 9(6):1–12, 06 2014.

[32] G. R. Kiran Kumar and M. Ramasubba Reddy. Filter bank extensions for subject non-specific ssvep based bcis. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 627–630, March 2019.

[33] Guangyu Bin, Xiaorong Gao, Yijun Wang, Yun Li, Bo Hong, and Shangkai Gao. A high-speed BCI based on code modulation VEP. *J. Neural Eng.*, 8(2):025015, 2011.

[34] Hirokazu Tanaka, Takusige Katura, and Hiroki Sato. Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data. *NeuroImage*, 64(Supplement C):308 – 327, 2013.

[35] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu. A speedy hybrid BCI spelling approach combining P300 and SSVEP. *IEEE Trans. Biomed. Eng.*, 61(2):473–483, Feb 2014.

[36] Jie Pan, Xiaorong Gao, Fang Duan, Zheng Yan, and Shangkai Gao. Enhancing the classification accuracy of steady-state visual evoked potential-based brain-computer interfaces using phase constrained canonical correlation analysis. *J. Neural Eng.*, 8(3):036027, 2011.

[37] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[38] Benjamin Wittevrongel and Marc M. Van Hulle. Frequency- and phase encoded ssvep using spatiotemporal beamforming. *PLOS ONE*, 11(8):1–18, 08 2016.

[39] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, 16(3):031001, apr 2019.

[40] Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of Neural Engineering*, 15(6), 2018.

[41] Feng Zhou and Fernando Torre. Canonical time warping for alignment of human behavior. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 2286–2294. Curran Associates, Inc., 2009.

[42] Chi Man Wong, Feng Wan, Boyu Wang, Ze Wang, Wenya Nan, Ka Fai Lao, Peng Un Mak, Mang I Vai, and Agostinho Rosa. Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs. *Journal of Neural Engineering*, 17(1):016026, jan 2020.